Dutch Stress Acquisition:
OT and Connectionist Approaches

Marc Joanisse[*] and Suzanne Curtin
USC Linguistics Department

1. Introduction: Theoretical Approaches to Phonology

The task of acquiring phonology is not a simple one; a language's phonological system involves a relatively complex set of patterns made up of abstract grammatical elements such as phonemes, moras, feet and prosodic words. And yet, children are able to acquire a great deal of their language's phonological system relatively quickly and efficiently. How are children able to acquire the sound patterns of language? The generative approach, which dominated much of the early thinking about language acquisition, suggests that children are born with an innate Universal Grammar (UG) which governs the phonological patterns of languages, and allows language learners to quickly and efficiently acquire seemingly complex linguistic systems (Chomsky 1965).

Earlier generative theories of phonology specified phonology as the knowledge of words' constituents (i.e., features and phonemes), and the types of rules that can act upon them (Chomsky and Halle 1968, Smith 1973). More modern theories have recast these ideas within a variety of other frameworks, for example Optimality Theory (Prince & Smolensky 1993), as well as expanding the domain of phonological constituency to include increasingly more abstract units, as in autosegmental phonology (Goldsmith 1990) and feature geometry (Clements 1985, Sagey 1986). Nevertheless, these approaches are faithful to many of the original theories in the generative tradition, such as the assumption that children are born with a relatively extensive amount of phonological knowledge. One principal tenet that is particularly relevant to the present work is the description of grammars as sets of deterministic symbolic mechanisms which act upon a discrete inventory of language tokens. In the case of phonology, generative grammars are described as rules, principles or constraints that act upon the phonemes and words encoded in the mental lexicon.

The present work reconsiders generative theories by exploring a radically different approach to how phonology is acquired, one in which children are not born with such a rich set of linguistic knowledge. This work instead considers the possibility that the input that children are exposed to contains a relatively rich set of information, and that children are able to use this information in order to acquire complex linguistic behaviors. This account does not reject the possibility that human language processing involves complex abstract structures such as features, phonemes, syllables, tones, feet and prosodic words. Instead, it suggests that the linguistic input to children is sufficiently rich as to allow them to acquire and use these concepts without the help of a great deal of innate linguistic structure. However, this approach does not include the core mechanics of generative grammars, namely the sets explicit and deterministic rules or constraints, and a lexicon that exists in isolation from such mechanisms.

1.1. Connectionist Phonology

In this work we develop an alternative view of phonology, in which phonological mechanisms are explained in terms of physiological factors such as perception

---

[*] We are grateful to Mark Seidenberg, Rachel Walker and the attendees of SWOT V for their useful comments on this work. Please address correspondence to Marc Joanisse, USC Linguistics GFS 301, Los Angeles CA, 90089-1693 email: marcj@gizmo.usc.edu.

and articulation, and cognitive constraints on learning and processing. This type of functional account is not new (Liljencrants & Lindblom 1972, Lindblom *et al*. 1984, Stampe 1979) though is has been highly controversial, often because of the difficulty involved in actually testing these theories. More recently, there has been renewed interest in functional theories, thanks in part to a better understanding of articulatory, acoustic and cognitive processing in both children and adults. The result has been a greater number of works which appeal to functional factors in explaining phonological patterns (among many others, Boersma 1998, Browman & Goldstein 1992, Flemming 1995, Hayes 1997, Steriade 1994, Stevens 1989).

Connectionist Phonology (CP) appeals to the functionalist approach by merging it with the idea that language is represented within a neural system, and as such is best characterized as the result of the basic principles which govern such systems (Joanisse 1999). The connectionist ("neural network") approach has emerged in the past two decades as an way of exploring how neural systems learn and represent cognitive processes (Rumelhart. & McClelland 1986, Elman *et al.* 1996) and in particular, language (see Seidenberg 1997 for a review). Connectionist networks provide a formal mechanism in which the influence of phonetic and auditory factors on phonological systems can be implemented and studied *in vitro*.

This approach has several benefits for studying phonology. First, while many behaviors can be described as rule-governed, many phonological phenomena are only partially regular. These include Japanese OCP effects (Ito & Mester 1986), English past tense verbs (Bybee & Slobin 1982), and Italian infinitives (Albright 1998). While generative principles can be used to account for the regularity in these types of systems, they have more difficulty accounting for the exceptional cases in which the rule is ignored. Connectionist models, in contrast, are ideally suited to explaining such quasiregular behavior, since they are able to encode both deterministic rules and their exceptions within a single architecture. This has been demonstrated in-depth for the case of English past tense verbs, which exhibit both a high degree of regularity (*raved*, *baked*, *lived*), but also a number of exceptional cases (*had, took, gave*). Implementing the English past tense in connectionist models has given researchers new insights into how morphological systems are acquired and used (Rumelhart & McClelland 1986, MacWhinney & Leinbach 1991, Joanisse & Seidenberg 1999). Likewise, we suggest that the CP theory might help in our understanding of quasiregular phonological alternations.

A second reason to use connectionist models to study phonological systems concerns the way in which they represent constraints, a popular topic in phonological theory thanks to the great deal of interest in Optimality Theory (OT, Prince & Smolensky 1993). The apparent contributions of OT, vis-a-vis constraints, are the notions of *universality* and *ranking*. All languages are proposed to use the same (innate) set of constraints, and differ only in how these constraints are ranked relative to one another. OT also specifies how ranking works, though a system of strict domination in which forms violating a higher-ranked constraint are always dispreferred over forms violating a lower-ranked constraint. This ranking system is itself considered universal and innate.

The current paper uses the CP framework to investigate a different set of issues related to phonological processes and constraints. Specifically, it examines whether such notions of constraint ranking as strict domination are sufficient to describe phonological phenomena, particularly with respect to phonological acquisition in children, and whether rule- or constraints-based grammars are capable of accounting for irregular linguistic processes. As we shall see, the results suggest that connectionist models can lend new insights to how linguistic grammars are best described, and how learning tends to proceed in such systems.

Dutch stress assignment is an ideal test case for the CP framework because it involves a great deal of abstract prosodic constituents, and because of the irregularity of Dutch stress that further complicates learning the regular pattern. Both these facts raise important issues of language learnability. Abstract phonological units cannot be directly observed and need to be inferred by a learner; likewise, the language learner does not have information about which words are regular or irregular at their disposal, since it is never expose to overt evidence to this effect. Both these facts greatly complicate the task of inferring a linguistic rule; it is proposed that the CP model might lend a better understanding of how children are able to learn phonological patterns. Finally, as we explain below, Dutch children seem to acquire stress in a way that suggests they are acquiring these patterns in a specific stage-like way. Accounting for these facts represents a further challenge to a CP model, which must accurately simulate how prosodic acquisition proceeds in children.

2. Dutch Stress Acquisition: Empirical Evidence

Stage-like development is a common trait of children acquiring language. For example, English speaking children acquire irregular morphological forms in what is often described as a U-shaped learning curve Brown 1973, committing overregularization errors (e.g., *TAKED) on forms they have previously produced correctly, and then apparently re-learning these forms. It has been argued that such behavior underlines the fact that children do not learn language through imitation, but are instead employing innate language acquisition mechanisms to acquire linguistic rules (Marcus *et al.*, 1992). Phonological acquisition in Dutch represents a similar type of case, which the present work examines by focusing on the errors children commit as they acquire main word stress in Dutch. These errors might reveal important facts about the kinds of learning mechanisms that are responsible for learning the phonological systems of language.

Dutch stress patterns are somewhat complex, but can be summarized as follows. Stress in Dutch is quantity-sensitive, meaning that heavy (-VV and -VC) and superheavy (-VVC and -VCC) syllables attract stress. In addition, main word stress tends to fall toward the rightmost edges of a word, usually on one of the last three syllables. There are several other important generalizations about stress in Dutch (Booij 1995, Kager 1989, van der Hulst 1984):

(1) Syllables containing schwas are not stressed, and stress is typically assigned to the left of a schwa-syllable.
(2) The antepenult cannot typically be stressed if the penult isclosed (-VC) or contains a diphthong.
(3) Words with final superheavy syllables or diphthongs have final stress.
(4) Words with open final syllables have penultimate stress.

Several different accounts have been developed to explain these facts, though they all call upon suprasegmental units, notions of syllabic weight, and generalizations about trochaic languages (of which Dutch is an example, Hayes 1991). However, these explanations leave out a few important facts. First, many Dutch words do not fit this set of rules, and instead have irregular stress. For example the city name *Amsterdam* receives final stress /amstərˈdam/, even though most other trisyllabic words with a penultimate schwa receive antepenultimate (initial) stress. Irregular stress is not limited to any single class of words (loan words, city names, words ending in “-dam”), indicating that these types of words have no special status in the language beyond their irregular stress patterns.

2.1 Stress in Dutch Children

Fikkert (1994) investigated the acquisition of stress in 12 Dutch children, ages 1;0 – 2;11. Her study used a longitudinal design in which words of various lengths and stress patterns were elicited from children in a somewhat structured environment. By recording each child's speech in 2-week intervals, over the course of several months, Fikkert acquired a large corpus of Dutch children's utterances. A primary finding of her research was the observation that all the children in her sample seemed to be producing similar types of errors as they acquired words. These errors also seemed to follow a stage-like pattern, such that children tended to produce clusters of one error type at a given point of development. Fikkert described these stages as follows:[1]

(5) Stages of acquisition in Dutch

Stage 1: At an early stage in acquisition, children were unable to produce polysyllabic words, and instead tended to truncate words. This typically involved producing only the final syllable (e.g., /balˈlon/ (*balloon*) → [ˈlɔn]).

Stage 2: As children began producing two- and three-syllable words, they tended to stress the initial syllable of a word, including words receiving non-initial stress (e.g., /balˈlon/ → [ˈbo:mɪ]).

Stage 3: Words were produced with level stress, such that more than one syllable seemed to be receiving main word stress (e.g., /balˈlon/ → [ˈbanˈdɔn]).

Stage 4: Children used adult-like stress patterns, though some phonemic errors occasionally still occurred (e.g., /balˈlon/ → [ba:ˈlɔn]).

There are two interesting nuances to these observations. First, children at Stage 3 do not produce every syllable in longer words with equal stress. Instead, Fikkert observed children dividing polysyllabic words into two equally stressed Feet, and assigning stress accordingly: /bu:rdəˈrɛi/ → [ˈbɔjəˈjɛi] (*farm*). Second, children still produced some stress errors at Stage 4, exhibited by Stage 2-type errors in words with non-initial stress. For example, one child at Stage 4 produced the word /ka:pi:ˈtɛin/ (*captain*) as [ˈpa:pitɛin].

It is clear that Dutch children do not produce these types of errors through simple imitation, since they are not likely to be exposed to words with stress on the wrong syllable, and they almost certainly have never heard words with main word stress on two different syllables. Instead, the child's errors seem to follow an apparently non-goal-directed pattern that suggests deeper principles are at play. Fikkert's account of this stage-like behavior is that it is the result of the child acquiring the use of abstract phonological knowledge. On her account, Dutch children call upon an increasingly complex set of prosodic units as they acquire stress, following the prosodic hierarchy first proposed by Selkirk (1980). Fikkert suggests that children begin acquiring main word stress using a bare minimum $(CV)_\sigma$ template, and progress through a quantity insensitive binary foot $(\sigma\sigma)_F$ (Stage 3), until arriving at a final stage in which their word template is a single stress-bearing prosodic word containing several feet (Stage 4.) In short, this stage-like behavior is an overt manifestation of the child learning to use increasingly more sophisticated aspects of the prosodic hierarchy.

---

[1] Throughout this paper, we use the convention of stating adult forms in slashes (/ /) and child forms in square brackets ([ ]).

2.2. A Closer Look at the Data

While the facts in (1-4) seem to suggest a uniform set of stages that Dutch children progress through, a closer investigation of these data indicate this is not the case. First, it is observed that stages of acquisition tend to overlap within individual children. As Fikkert herself acknowledges, the children in her study produced errors from more than one stage during a single session. For example, Table 1 illustrates how stages overlap in one of these children. At 1;10.25 Catooje produces errors consistent with Stage 2 (ˈpipa) and Stage 3 (ˈbo:ˈna:n). This pattern seems fairly common, as all children in the study seem to produce errors from more than one stage during a single session. Table 1 also illustrates a second trend in these children, specifically the tendency to regress from later to earlier stages from one session to the next. For example, at 1;10.25 Catooje produced mostly Stage 3 utterances, but then produces words consistent with Stage 2 at 1;11.10.

Table 1: A sample of Catooje’s Stage 2 and Stage 3 errors (in Fikkert 1994).

age	*stage 2*	*stage 3*
1;10.11	ko:ˈnein→ ˈkɪna:	
1;10.11	xi:tˈar → ˈhi:ta:	
1;10.11	ˈo:li:fant → ˈo:ma:	
1;10.25	paˈpir → ˈpipa	
1;10.25		baˈlɔn → ˈbo:ˈna:n
1;10.25		baˈlɔn → ˈbo:ˈnœn
1;10.25		ko:ˈnɛin → ˈkoˈnɛin
1;10.25		ko:ˈnɛin → ˈko:ˈnɛ
1;11.10	ba:ˈna:n → ˈbɪja:n	
1;11.10	ˈo:li:fant → ˈoŋam	

In addition, Fikkert does not analyze any of the child data with respect to irregular forms, so it is difficult to determine whether the stagelike behavior that children exhibit for extends to irregular forms, and whether children exhibit different patterns of acquisition for such forms. While these facts are not discussed in depth in Fikkert (1994), we suggest that any account of Dutch stress acquisition should be able to account for them, rather than being based on an idealization of the data. In the next section we review two types of generative accounts of Dutch stress acquisition, with regards to the degree in which they can account for the broader, less clear-cut data we have raised.

2.3. Generative Accounts of Stages

Fikkert calls upon a Parameter Setting learning theory in her account of Dutch stress acquisition (Dresher & Kaye 1990). In this theory, the space of possible phonological systems is limited by the parameters governing languages’ stress systems. This places strong limits the possible hypotheses that a language learner must posit, and thus guides how the learner searches through the space of possible languages (Clark 1992). While theories vary in their exact characterization of these parameters, there is some general agreement that they include parameters targeting the headedness of the prosodic word, foot shape, foot construction and defooting (for discussion, see Hammond 1990).

Parameter setting is not the only type of formal theory that could account for how children are learning Dutch stress the way they do. Here, we present an OT alternative, which illustrates how constraint ranking and re-ranking might account for the stage-like way in which Dutch stress is acquired. The general learning procedure in OT involves algorithmically re-ranking constraints based on surface forms perceived in the environment, using such methods as constraint ranking and demotion (Tesar & Smolensky 1996). Different behaviors emerge as these ranking change from their unmarked (default) to marked rankings. We will limit the OT account to the final three Stages, since Stage 1 (“one syllable”) is not informative with respect to stress assignment in Dutch children.

Stage 2 'One Foot'

The child's productions at Stage 2 are restricted to a foot. We will not; however, be discussing the content of the truncations (see Curtin, this volume, for an analysis of the truncation patterns). The stress pattern produced at this stage is consistently initial regardless of the adult pattern.

Table 2: The data for stage 2.

name	*adult form*	*child form*	*gloss*	*age*
Robin	ˌbu:rdəˈrɛi	[ˈbu:jɛi]	*'farm'*	1;10.21
	bɑˈlon	[ˈbydɔn]	*'balloon'*	2:1.7
	xiˈtar	[ˈsitau]	*'guitar'*	2;1.26
Catooje	koˈnein	[ˈkɪna]	*'rabbit'*	1;10.25
	paˈpir	[ˈpipa]	*'paper'*	1;10.25
Elke	ˈpɔrty:χal	[ˈpɔχɑl]	*'Portugal'*	2;3.27
	ˈo:li:ˌfɑnt	[ˈo:fɑnt]	*'elephant'*	
Noortje	ˌpɑnˈtɔfəls	[ˈtɔfhi:s]	*'slippers'*	2;8.29

At this stage, the child parses syllables into feet. The size restriction at this time is one foot, thus one stress is assigned to a left-headed foot. The necessary prosodic constraints are (Prince and Smolensky 1993):

(6) PARSE-σ: Syllables must be parsed into feet.
FTBIN: Feet are binary at some level of analysis.
WEIGHT-TO-STRESS (W-TO-S): Heavy syllables are stressed.

In the first example, the input is a bisyllabic word, the child produces a single foot with initial stress.

(7)

	koˈnein	PARSE-σ	FTBIN	W-TO-S
a	ko(ˈnɛin)	*!		
b	(ˈko)(ˈnɛin)		*!	
c	(ˈko)nɛin	*(!)	*!	*
d ☞	ˈ(konɛin)			*

Candidates (a) and (c) violate the PARSE-σ constraint while (b) violates FTBIN (as does (c)). The winner (d) only incurs a single violation of the lower ranked W-TO-S constraint.

Stage 3 'Two Feet'

The child is no longer truncating his/her productions at Stage 3. The child produces 2 feet and both of these feet are equally emphasized. This is due to the high ranking of PARSE-σ and the *demotion* of FTBIN (Smolensky 1996).

Table 3: Stage 3 data

name	*adult form*	*child form*	*gloss*	*age*
Robin	bɑˈlɔn	ˈbɑnˈdɔn	*'balloon'*	2;1.26
	ˈo:li:ˌfɑnt	ˈo:li:ˌˈfɑnt	*'elephant'*	2;4.8
Catootje	bɑˈlɔn	ˈbo:ˈna:n	*'balloon'*	1;10.25
	ˌte:ləˈfo:n	ˈhi:nəˈho:mə	*'telephone'*	2;0.6
	koˈnɛin	ˈkoˈnɛin	*'rabbit'*	1;10.25

The ranking for this stage is Parse-σ>>W-to-S>>FtBin.

(8)

	koˈnein	Parse-σ	W-to-S	FtBin
a	ko(ˈnɛin)	*!		
b ☞	(ˈko)(ˈnɛin)			*
c	(ˈko)nɛin	*!	*	*
d	ˈ(konɛin)		*!	

Candidates (a) and (c) lose on Parse-σ and (d) loses now loses due to the fatal violation incurred on W-to-S. The winning candidate (b) satisfies the high ranked Parse-σ constraint and the output is thus an equally emphasized foot. The same ranking applies for tri-syllabic and longer words.

Stage 4 'Adult-like'

The child, at stage 4, produces adult-like stress patterns. Notice however that there is still a tendency to produce initial stress for words like ‘*telephone*’.

Table 4: The data for stage 4

name	*adult form*	*child form*	*gloss*	*age*
Robin	ˌte:ləˈfo:n	ˈte:ləˌfo:m	*'telephone'*	2;1.7
	ˈo:li:ˌfɑnt	ˈo:vi:ˌɑnt	*'elephant'*	2;4.8
Tirza	koˈnɛin	ˌkɔˈnɛin	*'rabbit'*	2;2.1
	ˈo:li:ˌfɑnt	ˈɔli:ˌfɑnt	*'elephant'*	2;6.12

The language allows syllables to be licensed by the Prosodic Word. Due to this, Parse-σ is demoted. The winning candidate is thus (a).

(0)

	koˈnein	FtBin	W-to-S	Parse-σ
a ☞	ko(ˈnɛin)			*
b	(ˈko)(ˈnɛin)	*!		
c	(ˈko)nɛin	*!	*	*
d	ˈ(konɛin)		*!	

At this stage, the child has now obtained an 'adult-like' knowledge of the main stress assignment in Dutch.

To summarize, progression through the stages of stress acquisition involve the demotion of constraints (Smolensky 1996). As constraints are demoted, a new ranking gives rise to different optimal outputs.

Table 4: Summary of constraint re-ranking for stages 2-4 of stress acquisition.

Stage 2 'One Foot'	*Stage 3 'Two Feet'*	*Stage 4 'Adult-like'*
PARSE-σ FT BIN ↓ WEIGHT-TO-STRESS	PARSE-σ FT BIN ↓ WEIGHT-TO-STRESS ↓ FT BIN	Parse-σ ↓ WEIGHT-TO-STRESS ↓ FT BIN ↓ Parse-σ

While this provides an account of the regular stress pattern in clear-cut stages, there remain several questions. For example, what about acquiring the irregular stress patterns? In addition, there is a question of the variation within and across stages. Variation within a stage can be accounted for by constraints being unranked with one another (via partial ordering (Anttila & Cho 1998) or perhaps via multiple constraint hierarchies). This would account for overlap in the form of stage 2 and stage 3 productions co-occurring within a particular timeslot. However, variation in the form of regressive stages is not as clear cut (i.e. producing a stage 3 token when the stage 4 ranking has been established). In order to determine which direction to take within the OT framework, we implemented the CP approach.

2.4 Critique of OT93 and Parameter Setting accounts

The motivation for pursuing a connectionist approach comes from the apparent difficulties that the Parameter Setting and OT have in accounting for the empirical facts about Dutch, discussed herein. It is clearly difficult to critique OT as a single entity, since not every account adheres to the same principles governing what the universal constraint set is and what types of rankings are allowable. For example, many circumvent the problem of strict-domination ranking by allowing for such esoteric constraint interactions as local conjunction and multiple violations (Ito & Mester 1996). This allows for much more complex interactions of constraints, which in turn strongly influences the strength of OT in accounting for various behaviors. Instead, our critique focuses on OT as it was originally proposed in Prince & Smolensky (1993), or OT93. Since it is clear that, given some modifications, an OT account could be made sufficiently flexible to account for a broad range of data, part of our goal is instead to suggest what types of changes would be necessary for OT93 to account for Dutch stress patterns, and Dutch stress acquisition.

The first difficulty that generative accounts might have with these data relates to how they treat grammars and lexicons as separate entities. Because of this, it is difficult to reconcile the overlap between stages within the Parameter setting and OT93 frameworks. These theories contend that children's grammars are being adjusted independently of their acquisition of lexical items. Such theories would therefore predict that any change to the grammar – for example, re-ranking a constraint like PARSE-σ – should have a uniform effect on all words that the child knows. However, stages of language acquisition do not tend to be this crisp, as Dutch acquisition data indicates.

In some ways, an OT93 account might be compatible with these facts. For example, it is possible that these errors represent points at which the actual ranking of constraints is indeterminate. Thus, if the relative ranking of two critical constraints is not yet established, multiple outputs would be expected. Less plausible perhaps, are cases in which transitions between stages involve demoting an already ranked constraint. In such cases, the oscillation between stages would have to be the result of “un-demoting” and “re-demoting” a constraint. However, this type of behavior is not predicted by the standard Tesar & Smolensky (1996) constraint ranking learning algorithm, which only admits constraint demotion and would therefore not produce such a behavior. Moreover, it is unclear why this behavior would need to occur in this learner mechanism at all, given that this learning model does not require more than one exposure to a critical form in order to discover the correct ranking. This behavior is usually touted as a major benefit of this learning theory. As mentioned earlier, the second complication that generative accounts must contend with is the fact that not every word in Dutch has regular stress. Since Dutch children are exposed to both regular and irregular forms, they do not have information that would help them discriminate regular cases from the exceptions. This then raises the question of how generative learners are able to learn deterministic principles like the ones stated in (1-4).

Linguists have proposed two solutions to this problem. On the first account, irregular cases are treated as outside the scope of rules, and are instead explicitly memorized (Aronoff 1976, Bauer 1983, Pinker & Prince 1988, Pinker 1991). This allows the remaining regular cases to be accounted for using a relatively simple set of deterministic principles (e.g., rules, parameters, constraints.) The second way of accounting for such data is to incorporate both regular and irregular cases within the same framework. This has been attempted using elaborate sets of rules that vary in their specificity (Albright 1998, Chomsky & Halle 1968, Halle. & Mohanan85). However, the connectionist framework suggests a different alternative: rules and their analogues are in fact not the ideal way of understanding linguistic processes. Instead, cognitive behavior is best understood from the subsymbolic perspective, where probabilistic factors such as similarity and frequency interact in complex ways (Rumelhart & McClelland 1986, Seidenberg. & MacDonald 1999).

This type of framework has a particularly relevant benefit here: connectionist models of language do not treat the lexicon and grammar as separate entities, and so the derivational principle that a grammar acts upon the set of lexical entries in a language is not an issue in these models. Different “rules” can apply to different words, because information is encoded probabilistically and not as a specific set of rules. For the case of Dutch stress acquisition this means that, as the network learns the principles underlying Dutch stress, it can potentially commit different types of errors at a single point in training. This is because the model’s output for any given word is influenced by a number of factors, including frequency, phonological complexity (for instance, the number of syllables and phonemes) and similarity to other known forms.

3. The CP Model of Dutch Stress

3.1. Model overview

To simulate how Dutch children learn to produce stress, we trained a connectionist model to produce actual Dutch word forms, as sequences of consecutive syllables. This was done by presenting a recurrent network with a set of phonemes as input (e.g., /ˈo:li:fant/), and training it to output each phoneme in turn over a series of discrete time steps (e.g., [ˈo:] - [li:] - [fant]). The details of how this was done are further explained below. First, the motivations for using this particular model architecture should be explained. Because this model was not intended to represent a complete account of phonological acquisition, other types of linguistic tasks (such as word recognition and sentence production) were not included in this implementation. Instead, the current model instead focuses

only on processes related to word production, since the data we are addressing deals specifically with children's productions of words. The model architecture is presented in
Figure 1.

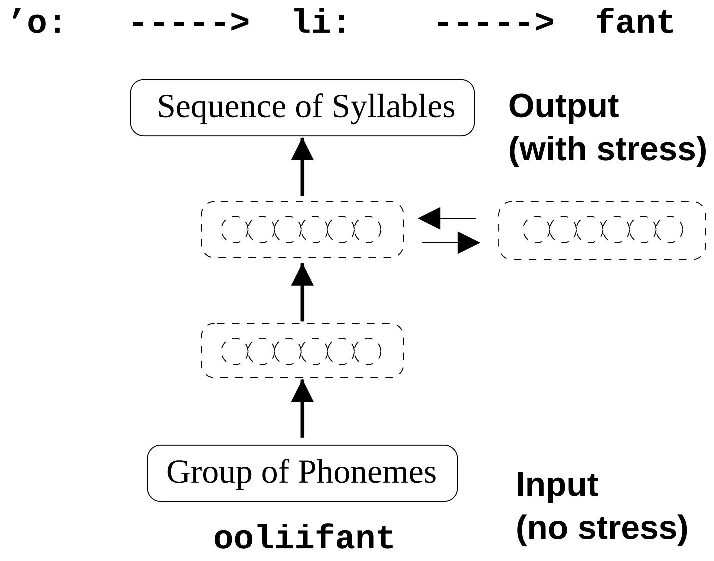

Figure 1: Network used to simulate word production in Dutch. The model was presented with groups of phonemes as input, and learned to produce the word as a sequence of syllables on the output.

The model received an input that represented a word that had to be produced on the output. Input words were represented as groups of phonemes that made up that specific word; for example, the word *ballon* ("balloon") would be /ballɔn/. Each phoneme was itself represented as a vector of 18 binary features corresponding to distinctive phonological features.[2] The input layer consisted of 19 single-phoneme "slots", and each word was left-aligned with these slots such that the first slot always corresponded to the first phoneme in the word, and so on. For example, the word "navigator" would be represented as naaviiɣaatɔr _ _ _ _ _ _ _, where underscores represent empty phoneme slots, encoded as a sequence of 18 0's. The choices of using discrete phonemes, as well as the specific features that were used, should not be interpreted as a commitment to a specific phonological framework on our part. The representations that were used merely reflect the need for a consistent mechanism for encoding the phonological content of words in a way that reflects the general degrees of similarity between different phonemes. It is proposed that any reasonable representational scheme should yield affected similar results to the ones presented here, given that the facts we wish to account for are only minimally related to issues of segments and features.

The model's task was to determine the word's syllabic structure and stress pattern based only on its phonological structure. For this reason, it did not receive any information about either of these on the input. Instead, this information was to be produced as the network's output. The output layer consisted of a single CCVVCC syllable "frame", made up of 6 phoneme slots identical to those on the

---

[2] voiced, voiceless, consonantal, vocalic, obstruent, sonorant, lateral, continuant, non-continuant, ATR, nasal, labial, coronal, anterior, high, distributed, dorsal, radical.

input, within which individual syllables of words could fit.[3] The model's recurrent architecture allowed it to produce sequential outputs. This in turn allowed us to model some of the dynamic aspects of word production, by training the network to produce sequences of syllables over several time steps. For example, the word *navigator* would be produced as the syllables [_naa_ _], [_vii_ _], [_ɣa _ _] and [_tɔ _r_], in that sequence. Words with fewer than four syllables were presented similarly, but were followed by the balance of empty syllables; for example *ballon* would be presented as [_ba_ _], [_lɔ_n_], [_ _ _ _ _ _] and [_ _ _ _ _ _]. Stress was assigned to a syllable by activating a single node, which was used solely for this purpose. For unstressed syllables, the stress node was set to 0 (inactive). An important feature of this system was that the syllabic position of every given phoneme on the output was not predictable from its position on the input. As a result, the task of determining the syllabic position of each of the input phonemes was not straightforward, and required the network to acquire important syllabification principles.

3.2. Training set

The training set consisted of Dutch nouns 1 to 4 syllables long drawn from the CELEX corpus of Dutch word forms (Baayen, van Rijn & Piepenbrock 1993). Nouns were used for two reasons. First, Dutch stress is most predictable for nouns, and so accounts of it have tended to focus on nouns (Booij 1995,Kager 1989, van der Hulst 1984). Second, the Dutch acquisition data we wish to account for also deals exclusively with nouns, which is perhaps not surprising given that children tend to acquire mostly nouns before the age of 2 (e.g., Macnamara 1972, Gentner 1982). CELEX lists 33,553 4-syllable nouns. However, hardware limitations made including all these forms in network training impractical. For that reason, the training set was reduced by randomly selecting 10% of these words. In addition, a number of homophones were removed, leaving us with 3,324 Dutch nouns that were used in the training set. Stress assignment and syllabification were also obtained from CELEX, and was based on CELEX coders' judgements.

Closer investigation of this corpus revealed several statistical characteristics worth noting here. First, there is an inverse relationship between words' token frequencies and the proportion of words in the training set in a specific frequency range, such that there are many more low frequency words compared to high frequency words, a trend that not unexpected (Zipf 1935). Second, monosyllables seem to be an exception to this trend, witnessed by the greater proportion of higher-frequency monosyllables in Dutch. Both these trends are also present in the training set, indicating that the network was exposed to this statistical information. Finally, although monosyllables tend to have higher token frequencies in Dutch, the language as a whole has a relatively small proportion of monosyllabic words, compared to polysyllabic words. This fact was also captured in the training set, and could also have an impact on how Dutch words are learned.

3.3. Training procedure

At the beginning of training, all network connection weights were randomized between 0.01 and -0.01. Training proceeded as follows: at the beginning of each

---

[3] To reduce the size of the training set, and to shorten training time, the model was not trained on words with CCC clusters in either the onset or coda position. In practice, words of this type consisted of less than 7% of all Dutch words, and of these, only 3 words had CCC clusters in the coda position. This would suggest the impact of omitting these words from the training corpus was minimal.

training trial, a word was selected from the training set. Selection was frequency weighted such that the probability of selecting a given word was a function of the word's logarithmic frequency in the CELEX corpus. (Using a log frequency transform assured us that the network would receive a sufficient number of exposures to low frequency words within a reasonable number of training trials.) Activation was allowed to propagate throughout the network for 10 time steps, after which connection weights were adjusted for each of these time steps, using the backpropagation through time learning algorithm Williams & Peng 1990) and cross-entropy error calculation. A logistic activation function was used; the learning rate was set to 0.001; error tolerance was set to .1, meaning that error correction was not applied to activations within this level of tolerance.

4. Training Results

Network training was stopped after 5 million training trials. Performance was assessed by presenting the network with words in the training set and comparing each phoneme in the resulting network output to the correct output. A nearest-neighbor method with a Euclidean distance function was used for this purpose. Errors were registered when the network produced an output that most closely matched a Dutch phoneme that was not the intended output (e.g., producing [g] instead of [ɣ].) For our purposes, an entire word was coded as phonologically incorrect if any of the syllables of a word contained an incorrect phoneme. Errors were also registered for cases in which the network failed to produce a phoneme at all, or it produced a phoneme where none was expected.

Stress assignment was assessed by directly measuring the activation of the stress node in the output layer while each syllable was output. A syllable was considered stressed if the stress node produced a value of greater than or equal to 0.5 during that syllable's time step, and unstressed if the stress node had an activation of less than 0.5. While we were most interested in the patterns of behavior that the model exhibited over time, our first set of analyses focused on qualifying the network's behavior at the end of training. This allowed us to assess both whether it had learned a significant portion of the training set, and whether the errors it was producing were consistent with those of children.

This was done by presenting each word in the training set to the model at the end of training. The network correctly produced the phonological forms of 89% of the training words, and predicted the correct stress for 94% of the training words. Below we present an overview of the types of errors the network was producing at the end of training.

4.1.Segmental Errors

The words that the network produced incorrect phonological forms for seemed to be ones that represent difficult cases, namely low frequency and longer words. The mean log frequency of the incorrect words was 1.25, indicating that most of the phonologically incorrect words had frequencies below 20 per 4.2 Million in the CELEX frequency counts. In addition, phonological errors tended to occur in longer words: only 22% of the phonologically incorrect words were 1- or 2-syllables long, while 37% and 41% were 3- or 4-syllable words, respectively.

Stress assignment was remarkably good by the end of training; only 6% of words in the training set were incorrectly stressed. Stress errors fell into three categories. In some cases, the network assigned stress to the incorrect syllable. Elsewhere, the network assigned stress to more than one syllable on the output. Both of these error types are of interest to us, because of how it might compare to children at Stages 2, 3 and 4 in Fikkert (1994). A third type of stress error also occurred, when the network was not able to determine which syllable was stressed; in these cases the stress node was below threshold for all syllables in the word. This error type is more difficult to compare to errors in child language, because it can be interpreted two ways. The first possibility is that the network is applying equal stress to each syllable, for whatever reason, and this is resulting in no syllable receiving full stress. The second possibility is that the network is attempting to apply greater stress to a specific syllable, but the stress node is not above threshold for that syllable. Overall, the patterns of stress errors that the model produced at the end of training seem consistent with the types of phonological errors committed by children ages 2 to 5. They suggest that, while performance had not yet reached an adult-like level of performance when training was stopped, the model's performance was similar to that of children who have acquired much of Dutch phonology.

5. Developmental Patterns in the Model

The primary issue we wished to address related to the patterns of behavior the network exhibited over the course of training. This was investigated by recording the network's weight-state every 100,000 training trials, and testing the network's performance on a variety of word types at these intervals. Testing the network over the course of development is analogous to a longitudinal study of an individual child, and allows us to compare the network's developmental patterns to that of children acquiring Dutch stress.

5.1. Regular and Irregular Stress

Before looking at specific developmental patterns in the network, we were interested in whether the network had indeed acquired *rule-like* behavior at all. It is conceivable that the network was doing nothing but "memorizing" the stress patterns of each word it was exposed to. Because Dutch words vary in the regularity of their stress patterns, it was possible to assess whether this was in fact how the model acquired stress. This was done by comparing the network's performance on phonologically similar words with regular and irregular stress patterns, over the course of training. If the model were simply memorizing stress patterns, we would not expect to see differences between the two types of patterns.

Since there is little empirical data comparing the acquisition of regular and irregular stressed words in Dutch, it is difficult to determine at this point whether these results are accounting for the way in which children learn such forms. However, there is at least one area of Fikkert's study that might shed some light on the effects of regularity on Dutch stress acquisition, related to stress errors in children at Stage 4. While the general trend at this stage is for children to produce words with correct stress, this was not always the case. Table 5 gives some examples of these errors, as listed in (Fikkert 1994). It is observed that these errors in children at Stage 4 seemed to be concentrated in low frequency words with irregular stress. These appear to be overregularization errors, because the final syllable is not the typical location for stress in words of these forms (van der Hulst 1984). This indicates that they have learned the regular stress pattern for such words, but that they are tending to apply it to words with irregular stress.

We investigated this effect in the connectionist model by testing it on 3-syllable words with irregular stress, at a point in training consistent with Stage 4 (after 3.5 Million training trials). Although the model was producing the correct stress for many irregular forms at this point, several initial-stress errors were observed on lower frequency irregularly stressed forms. These errors are reproduced in Table 6.

Table 5: Sample of errors on irregularly stressed words at Stage 4 (data from Fikkert 1994).

child's name	*age*	*adult form*	*child form*
Robin			
	2;4.29	amstərˈdam	→ ˈapstədam
	2;4.29	pa:ra:ˈply:	→ ˈpaləply:
	2;4.29	kro:ko:ˈdɪl	→ ˈko:kədɪw
Tirza			
	2;1.17	kro:ko:ˈdɪl	→ ˈko:kəltɔ
	2;3.27	amstərˈdam	→ ˈɛmstədɛm
	2;5.5	ko:nɪˈŋɪn	→ ˈko:nɪŋɪŋ
Enzo			
	2;2.4	kro:ko:ˈdɪl	→ ˈnɔkɔdɪl
	2;3.14	bu:dəˈrɛI	→ ˈbudərɛi

Table 6: Model performance on irregularly stressed forms at Stage 4.

target form	*model output*	*frequency (/ 4.2 Million)*
di:glo:ˈzi:	ˈdi:blo:zi:	0
ho:xerˈwal	ˈho:zərwal	0
sa:ti:ˈnɛt	sa:ˈti:nɛt	0
sy:pərˈfly:	ˈsy:pərpli:	3
tri:jarˈxi:	tri:ˈjarxiy	0

These errors are significant because they further suggest that the model was producing rule-like behaviors consistent with what is observed in actual children. Crucially, while the model was capable of encoding most irregular forms, it was not accomplishing this solely by memorizing the training corpus; instead it learned important generalizations about how stress is applied, and committed errors that are consistent with these generalizations. And while these types of overregularization errors are not a unique characteristic of connectionist models (Marcus *et al*. 1992), they do indicate that this type of learning mechanism can account for a wide variety of data related to Dutch stress acquisition.

While empirical studies have tended to ignore the effects of regularity on prosodic acquisition, the results of the present modeling work seems to suggests that this is an important avenue of future research in this domain. Indeed, this work makes strong predictions as to the importance of pattern regularity in the acquisition of phonological systems, and suggests that developmental profiles can extend well beyond well-researched U-shaped learning patterns.

5.2. Pools of Regularity.

The second aspect of irregular stress we were interested in concerned words with irregular but consistent stress patterns. For example, two-syllable words in Dutch of the form $(VC)_\sigma$ $(VC)_\sigma$ tend to have initial stress, since Dutch is a trochaic language. However, van der Hulst (1984) has observed that certain non-morphological word endings tend to correlate with final stress placement in such words, for example the "French" word endings illustrated in Table 7. Thus, it would appear that factors other than syllabic weight seem to influence the placement of stress in Dutch.

Table 7: Sample of Dutch VC forms that appear to attract final stress in bisyllabic VC-VC words.

word ending	*initial*	*final*
-et	4	30
-el	0	5
-on	6	16

The tendency for (non-morphological) segmental factors to affect stress patterns is relatively common in Dutch. In fact, it has been shown that the ability to learn Dutch stress is greatly enhanced when segmental information is taken into account. Dalemans *et al.* (1994) applied a similarity-based learning algorithm (a class of statistical learners similar to the type of neural networks used here) to learning Dutch stress. They found that the mechanism's ability to assign the correct stress to words was greatly enhanced when both segmental and metrical information was made available to the artificial learner, compared to when it was only given metrical information. A drawback of the Dalemans *et al.* learner was that it was not a model of development, and could not account for how children acquire stress in stage-like ways.

Because the network we used here also had both types of information at its disposal during training, it is possible that it too could account for these facts. We tested this by presenting it with words containing well-known segmental cues to stress. Since the network was only trained on a sample of the words in CELEX, we were able to test it on sets of VC-VC words in CELEX that were not in the training corpus. As Table 8 shows, the fully trained model did indeed tend to apply final stress to words ending in *-et -el* and *-on*. By way of comparison, we have also indicated the model's performance on various VC endings that seem to be highly correlated with initial stress (*-um*, *-aN* and *-is*). For these types of words, the model appears to be consistently applying initial stress. Together, these data indicate that the network has learned several types of information relevant to Dutch stress, not merely suprasegmental factors such as syllabic weight and canonical trochaic stress patterns.

Table 8: Network's performance on bisyllabic word forms with phonologically predictable final (upper) and initial (lower) stress.

pattern	*% correct*
VC-ˈ<et>	65
VC-ˈ<el>	75
VC-ˈ<on>	68
ˈVC-<um>	91
ˈVC-<aN>	100
ˈVC-<is>	79

These results are significant because they underline an important difference between the connectionist approach and symbolic learning theories; whereas the traditional generative accounts of Dutch stress have focused solely on factors such as syllable weight in deriving the rules of stress, it is clear that other types of factors are also relevant. In the present example, stress in words consisting of two heavy (VC) syllables should only be predictable from suprasegmental information, in a traditional metrical account. However, phonological information such as the coda of the word's final syllable seems to modulate these factors. The network we used took advantage of these types of subregularities in the course of learning the training corpus in order to better acquire the stress patterns of Dutch words, and this in turn also allowed it to generalize to unfamiliar irregular words.

This result underlines a broader difference between the type of learning mechanism we are proposing here and the one theorized by Pinker & Prince (1988), *et seq.* While the generative account has often proposed that regular patterns are acquired using linguistic rules, irregular patterns are strictly memorized, and are only very rarely extended to novel words. The current results make the strong prediction that Dutch speakers will tend to extend the irregular "French" stress pattern to other words with similar endings, contrary to what is predicted in a Pinker & Prince type model.

5.3. Stages of Acquisition in the Model

The second aspect of Dutch acquisition we were interested in concerned the stage-like way in which children appear to learn stress patterns. The network's behavior over the course of training was investigated to assess whether it was also producing the types of errors described by Fikkert (1994). This was done by presenting the network with the entire set of 2- and 3-syllable words in the training corpus, and comparing its output to the "adult" target output. A syllable was considered stressed if the activation of the stress node on the output was greater than 0.5. A Stage 2 error was coded when the network incorrectly produced an initial stress for a word. A Stage 3 error was coded when the network produced more than one stressed syllable (that is, the stress node was activated above threshold for more than one syllable). Examples of errors consistent with Stage 2 and 3 are listed in Table 9.

Table 9: Examples of errors produced by the model that were consistent with errors at Stages 2 and 3 in Fikkert (1994).

	word	*adult form*	*model output*
Stage 2			
	echec	e:ˈʃɛk	ˈe:sɛk
	donaat	do:ˈna:t	ˈdo:na:t
	jubee	jy:ˈbe:	ˈjy:be:
	Francaise	franˈsɛ:sə	ˈfransɛ:sə
	diffusor	dɪ ˈffy:zɔr	ˈdɪffy:zɔr
	woestijnrat	wu:sˈtɛɪnrat	ˈwu:steɪnrɛt
Stage 3			
	framboos	framˈbo:s	ˈframˈbo:s
	vignet	vɪnˈjɛt	ˈvɪnˈjɛt
	santon	saˈntɔ:	ˈsanˈtɔ:
	haverzak	ˈha:vərzak	ˈha:ˈvərzax
	huisorde	ˈhʊɪsɔrdə	ˈhʉɪsˈtɔrdə
	jeugdbeleid	ˈjøɣdbəlɛɪtd	ˈjøɣdbəˈrɛɪt

The proportion of errors consistent with Stage 2 and Stage 3 that were produced by the network is plotted in

Figure 2. These results indicate that the network was producing a significant number of both types of errors during training, and that the proportions of these kinds of errors changed as training progressed. Consistent with the child language data, the model was also producing many more Stage 2-type errors earlier in training, and many more Stage 3-type errors later in training. In addition though,

the behavior was not perfectly stage-like. Instead, there was a clear transition period during which the network produced a significant number of both error types.

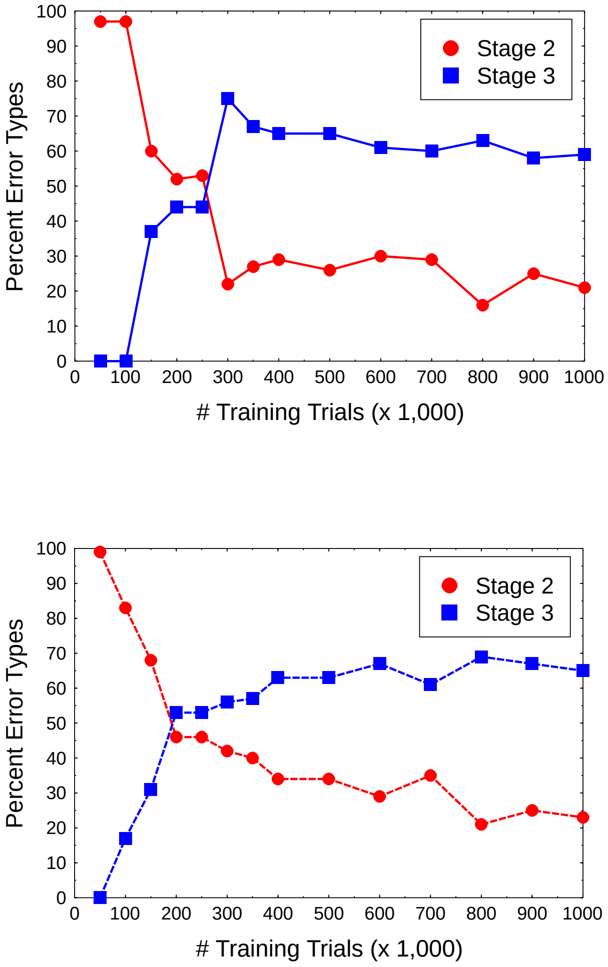

Figure 2: Proportion of Stage 2 and 3 errors on 2-syllable (Top) and 3-syllable (Bottom) words in the network, over the course of 1 million training trials.

Figure 3 illustrates how the Stage 3-type behavior began to decline later on in training, as the proportion of overall correctly stressed words reached asymptote. Here again, the network's performance seems to be consistent with the data from Dutch children, who also demonstrated fewer Stage 3 errors as they produced more correctly stressed words.

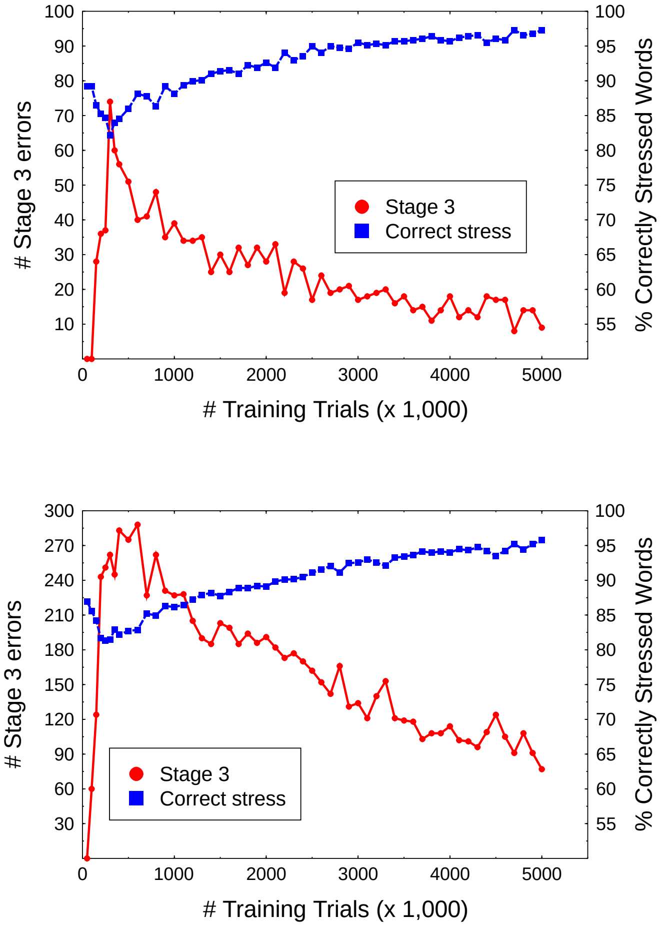

Figure 3: Proportion of Stage 3 errors in the network, relative to percent correctly stressed words, over the course of training. (Top: 2-syllable words. Bottom: 3-syllable words).

Given these results, it would appear that the network was learning to produce stressed words in a similar way to the Dutch children observed by Fikkert – it showed stage-like behavior with respect to the types of errors it produced, but at the same time there was a gradual transition between stages. As such, the network was producing appreciable numbers of errors consistent with Stage 2 and 3 at one point in training, and produced many errors consistent with Stage 3 at the same time as it was producing Stage 4-type behavior.

The source of the network's errors appears to be related to how it was learning the principles governing Dutch stress. Initially, the network was defaulting to initial stress on all words, because of the large proportion of initially-stressed words it was exposed to. Over the course of training, the network began producing Stage 3 errors as a result of developing the notion of weight-bearing syllables, and foot headedness. However because the network had not yet fully acquired the rule for placing stress on the correct syllable, it instead tended to stress two separate syllables in the word. These errors subsided as the network’s ability to correctly stress words of different lengths reached asymptote.

6. Discussion

This work has investigated various aspects of phonology, morphology and acquisition that have previously been addressed within the generative framework. The purpose was to examine how these data, and others like it, might be well suited to a connectionist framework in which linguistic patterns are learned and processed in a probabilistic mechanism. Such an account contrasts with generative theories in several important ways, as is discussed below. This discussion focuses in particular on OT-type approaches in order to address the similarities and differences between the CP approach and other constraints-based theories to phonology. As we shall see, the data from Dutch underline several important distinctions between the two types of theories.

6.1. Quasiregular Domains

A distinguishing aspect of connectionist models is how they encode linguistic grammars. While generative approaches treat grammars as specific sets of rules or constraints, the connectionist theory instead encodes them as generalizations drawn from regularities in a language which are then applied to unfamiliar and novel forms. An important advantages to this approach is its ability to encode cases which fail to conform to the regular pattern within the same mechanism as those which do follow the regular pattern. This is particularly helpful in instances where it is difficult to determine which cases are the “regular” ones. For example, Albright (1998) has examined aspects of Italian morphology in which several different patterns behave like the regular pattern (or, at the very least, fail to behave like irregulars) in specific contexts. For example, the Italian infinitive marker surfaces in different forms depending of the class of the verb it is affixed to (e.g., *sed-ˈare, ˈled-ere sed-ˈere, sped-ˈire*). The first ending, -*ˈare*, is typically described as the productive ending; that is, this ending is typically applied to nonce verbs; thus it is typically assumed to be the “default” or “rule,” and that all others are irregulars (and thus non-productive). However, all four verb classes tend to be internally consistent, such that class members tend to bear some resemblance to one another. Albright showed that native Italian speakers are sensitive to the neighborhoods of regularity formed by Italian verb classes, and as a result show preferences for the appropriate “irregular” ending for nonwords that are similar to other irregular verbs (e.g., speakers tended to prefer *adˈduŋg-ere* over the putative default *adduŋg-ˈare.*) Such cases are problematic for dual-mechanism type explanations that posit a single grammatical rule and a list of exceptions, since it is difficult to determine which instance actually represents a rule, and why the exceptional cases show such a high degree of productivity.

A similar problem arises in the case of Dutch, because of the degree to which its stress patterns deviate from the default pattern. As we have investigated here, there appear to be tradeoffs between different types of regular patterns such that the broader pattern of metrically driven stress is overridden by segmental factors (as in the case of words ending in *-et*). These so-called “pools of regularity” are easily explained in the CP framework, because of how such a learning mechanism allows for multiple forms of constraint, and because of how such mechanisms encode generalizations probabilistically, not as explicit rules.

In many ways, this type of account is also compatible with ideas from OT93 which allow for complex interactions of constraints within a single mechanism. On such an account, constraints deriving stress from a word’s metrical structure could be outranked by constraints which would force irregular stress patterns in words containing specific segments; given the sufficient constraints, the default regular pattern could be overruled by the irregular pattern in certain cases.

A second type of account for this type of behavior has also been proposed, in which different word classes seem to be targeted by different phonological processes *cophonologies* (Inkelas, Orghun & Zoll 1996, Ito, Mester & Padgett 1995). On this account, the lexicon is divided into predictable strata that encode specific classes of words, (such as loan words, native words, words with a specific metrical or segmental structure). Separate sets of constraint rankings coexist in the grammar for each lexical stratum. The result is different phonological processes targeting discrete sets of words. This type of system would allow for separate productive processes to occur in a single language, though it does not allow for what are called *static* processes, patterns that are not productive. As in other generative accounts, these patterns are simply memorized, and are not governed by grammatical principles.

We argue that neither type of OT system is sufficient to account for the broader range of facts about Dutch. First, not all words taking irregular stress do so predictably; words like *tonˈsil* and *verˈnis* have irregular stress, but do not appear to belong to a predictable class of other irregularly stressed Dutch words. Second, not all words with stress-attracting endings take irregular stress; for example the usual irregular (word-final) stress pattern seems to be overruled in some words (e.g., *ˈclicket, ˈTibet, ˈticket* have regular stress). For both types of words, any generative system would have to specify these words as memorized (or lexicalized) irregulars. An OT system that uses a single constraint ranking would incorrectly produce irregular stress on words like *sorˈbet* (a word that takes regular stress, but which is phonologically similar to the *-et* class of irregulars.) The cophonology solution might be more capable of accommodating such facts, by arbitrarily assigning each word in the lexicon to a specific cophonology without regard to how they might be organized. However, creating arbitrary groups of lexical items presents a special problem for this type of account, because it abandons the principle that lexical cophonologies are organized by some metric of similarity (be it semantic or phonological). In addition, such an account cannot explain the strong effect of similarity on the generalization of apparently static irregulars patterns to novel words. Inkelas et al. (1996) suggest such patterns are not suitable for a grammar-based account and should instead be memorized. However, this fails to explain why speakers prefer irregular ending in some nonwords (e.g., English irregular past tenses *spling-splang*). Albright's data on Italian infinitives is also informative in this regard (Albright 1998).

6.2. Rules and Lexicons

As discussed throughout this paper, a major advantage to the CP approach is the absence of either specific rules, or an explicit lexicon in which word-specific information is encoded. This means there is no need to posit a pattern as memorized or rule-governed; likewise, the Prasada & Pinker (1993) distinction between rule application (*rick-ed*) and irregular generalization (*splang*) is vacuous. In fact, the connectionist framework allows for a continuum between highly regular, and highly irregular linguistic patterns. For instance, the consistency of stress patterns varies cross-linguistically; Russian stress is lexical, meaning stress is assigned idiosyncratically and is thus highly unpredictable; English and Dutch have more regular stress patterns but also many exceptions (which can themselves be quasi-predictable, e.g., English bisyllabic verbs often take word-final stress); and Finnish and French have (almost) perfectly predictable stress. On the connectionist account, all these types of systems are simply points along a continuum in which generalization trades off with word-specificity, and all such systems can be encoded within a connectionist architecture.

The consistency of linguistic patterns also varies *within* languages. Morphological and phonological paradigms can consist of different patterns that generalize to different degrees. Dutch stress is an example of this - some, but not all, irregularly stressed words in Dutch form a semi-regular set (e.g., French-sounding words ending in *-et*), and there are also exceptions to these exceptions (e.g., ˈ*ticket*). As we discussed above, this is problematic for accounts using deterministic grammars that are discrete from the lexicon.

The connectionist account provides us with an alternative in which all degrees of regularity can be accounted for. It encodes the regularity of morphological and phonological patterns as a result of learning the words in a language. Because the network architecture is not sufficient to overtly memorize each form that the network comes across, the network must instead encode the regularities that it comes across in the input, including the circumstances under which specific regularities fail to occur. The range of statistical regularities that the network can encode is relatively unbound; the class of network used in the present simulation network is in theory capable of fitting a wide variety of statistical regularities that it is exposed to. It is only limited by the computational capacity imposed by such factors as the number of hidden units it has, and the nature of the input it is exposed to. This explains why it is not limited to a single source of constraint; segmental and metrical information are both available to the network, and both appear to be playing roles in its behavior. The application of one constraint over another is determined by the statistical probabilities that the network has inferred from the input it is exposed to.

6.3. Gradience of Acquisition

The present results suggest another benefit of the CP approach, which is the ability to account for a broad range of language acquisition data. Generative theories language acquisition suggest that stages of acquisition as the result of qualitative changes to the child’s grammar that act upon all words in their lexicon equally. However, the data from Fikkert (1994) suggest that this does not accurately describe how Dutch children actually learn phonology, given the gradience in children’s transitions from one stage of acquisition to the next. The present theory suggests such patterns result from how learning tends to proceed in neural systems, typically in a graded and smooth fashion. In addition, it provides an alternative explanation for the errors Dutch children produce at these stages of acquisition.

The model’s behavior over the course of learning was strongly influenced by characteristics of the training set that it was exposed to. Analyses of Dutch syllable and stress patterns revealed biases toward words with initial stress and words with one or two syllables. (While Dutch has fewer monosyllables than 2- or 3-syllable words, statistics from CELEX indicated that speakers are nevertheless exposed to many monosyllables due to their tendency toward higher token frequencies.) Errors at Stage 2 reflect the model’s response to this bias toward initial stress. Errors at Stage 3 reflect a greater sophistication on the part of the model as it attempts to fit the broader patterns in the training set, including the tendency for each foot to receive stress (e.g., [ˈtrɔmpɛsˈtest] for trɔmpɛtˈtIst). Errors at this stage also reflect competition between the Stage 2 tendency towards initial stress, and the need to uphold generalizations about moraic trochees (resulting in errors like [ˈsu:nda:ˈne:s] for su:nda:ˈne:s). Changes in the strategies the model is using to learn stress patterns is also similar to that of Dutch children. Learning is characterized as the gradual progression from one stage to the next, rather than stepwise shift from one type of behavior to the next.

6.4. Learning in CP and OT

The probabilistic learning that occurs in connectionist models seems to conflict with how generative grammar is typically thought to be acquired. In this section we focus on how probabilistic learning relates to OT, which has typically posited constraints as being strictly ranked relative to one another. The prevailing theory of learning in OT characterizes acquisition as the process of ranking and re-ranking these constraints based on the surface forms it is exposed to. We argue that the result is a system which cannot produce the types of probabilistic behaviors discussed in this work, and which has no clear mechanism for accounting for the range of quasiregularity that exists in languages.

This first issue relates to the system of strict domination that OT93 uses to encode grammars. Such as system uses a single set of constraints with a specific ranking to assess the optimality of all a language's words. However, the way in which Dutch is acquired casts some doubt on this theory; while we have shown that the individual stages with which Dutch is acquired can be captured using an OT93 grammar, it cannot completely account for the way in which these stages tend to overlap. The way in which learning in the OT93 model is proposed to occur (Tesar & Smolensky 1996) could allow for some periods of uncertainty when the rankings of crucial constraints are indeterminate, but it cannot explain the tendency toward regressions to earlier stages. It also predicts that between-stage transitions will be characterized by a random oscillation between the two types of behavior, which again does not seem to be borne out by the empirical data.

More recent developments in OT might be better suited to addressing these issues. Boersma & Hayes (1999) have proposed an alternative to the strict ranking principles in OT93 in which constraint rakings are proposed to be probabilistic rather than deterministic. The actual ranking of constraints are expressed as areas of probability, allowing relative rankings to be indeterminate to varying degrees. This allows actual constraint rankings to shift from one utterance to the next, but also bounds the degree to which these shifts can occur. The result are behaviors consistent with different types of utterances (e.g., stress errors consistent with Stages 2 and 3) emerging from a single grammar. A major benefit of this approach is the ability to capture the relative frequency with which different alternate behaviors tend to occur. In the case of Dutch, this might be used to account for why children progress gradually through stages. Learning in such a model is similar to other OT learning accounts, and involves making small incremental adjustments to constraint rankings.

Because Boersma & Hayes' model can account for the overlapping stages observed in Dutch stress acquisition, it clearly represents an important development in understanding the range of constraint interactions germane to phonological systems. Nevertheless, it does seem to fall short in accounting for the range of quasiregular behaviors observed in languages like Dutch. Like other OT models, it is difficult to reconcile this type of mechanism with the facts about irregulars in languages, since it seems to predict that productive and static phonological patterns are distinct, in spite of emerging evidence of gradience between these two extremes. This problem extends to acquisition, since any learning mechanism must be able to acquire grammatical generalizations in spite of the quasiregularity of linguistic patterns. We argue this difficulty remains problematic for all OT accounts, and deserves further attention from the point of view of formal phonology.

7. Conclusion

The goal of the current research was to explore how stress is acquired in Dutch, and to explain the stage-like way in which learning proceeds in Dutch-speaking children. The motives for using a connectionist model should not be misconstrued; the current work does not dispute the basic linguistic principles underlying stress (e.g., suprasegmental units such as moras and Feet). Indeed, it would appear that these types of principles represent crucial elements of many types of prosodic behavior, of which stress assignment is only one example. Nevertheless, the present work also suggests that many relevant types of information are present in the input to the child, and that the child's grammar need not explicitly encode all of the relevant information. In addition, it raises the possibility that a strictly symbolic (i.e., grammar-based) account of stress is inadequate to account for a variety of behaviors demonstrated by children acquiring language.

Part of our goal has been to address ways in which current constraints-based theories, which have shown promise in accounting for many "hard problems" in phonology, still appear to fall short in accounting for a range of data related to irregularity and stage-like acquisition. There seem to be two specific types of data that we propose remain difficult for OT accounts, in particular the OT93 approach. The first involves the fact that the different stages acquisition can overlap within children acquiring Dutch; these children produce errors of two stages in a single session, and regress to earlier stages from one session to the next. As we have shown, the notion of ranking and demoting constraints seems somewhat inconsistent with this type of behavior, and perhaps represents an oversimplification of the acquisition process. The second type of data is the case of irregularly phonological patterns, and how they are learned. While several different accounts have been put forward to handle these cases, it remains unclear whether an OT learning mechanism can acquire both regular and irregular linguistic patterns, and show the types of generalization behaviors often observed in native speakers. The connectionist account presented here suggests that a deeper understanding of the empirical data is necessary in order to better characterize both the task that the children confront as they learn quasiregular patterns, and the type of learning mechanism that can account for their ability to do so.

References

Albright, A. (1998). Phonological subregularities in productive and unproductive inflectional classes: Evidence from Italian. M.A. Thesis, UCLA, Los Angeles CA.

Anttila, A. & Cho, Y. (1998). Variation and change in Optimality Theory. *Lingua* 104, 31-56.

Aronoff, M. (1976). *Word formation in generative grammar* Cambridge MA: MIT Press.

Baayen, R.H., Piepenbrock, R. & van Rijn, H. (1993). *The CELEX lexical database (CD-ROM)*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Bauer, L. (1983). *English word-formation*. Cambridge, UK: Cambridge University Press.

Boersma, P. (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. PhD Thesis, University of Amsterdam.

Boersma, P. & Hayes, B. (1999). *A ranking algorithm for free variation*. Talk presented at 73rd annual meeting of the Linguistics Society of America.

Booij, G. (1995). *The phonology of Dutch*. Oxford UK: Oxford University Press.

Browman, C.P. & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica* 49 (3-4), 155-180.

Brown, R. (1973). *A first language: The early stages*. London: George Allen and Unwin.

Bybee, J. & Slobin, D. (1982). Rules and schemas in the development and use of the English past. *Language* 58, 265-289.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. & Halle, M. (1968). *The sound pattern of English.* MIT Press.

Clark, R. (1992).The selection of syntactic knowledge. *Language Acquisition* 2, 83-149.

Clements, G. (1985). The geometry of phonological features. *Phonology Yearbook* 2, 225-52.

Curtin, S. (this volume) Positional Prominence and Size Restrictors in Prosodic Development, Proceedings of the 5th Annual Southwest Optimality Theory Conference, La Jolla CA: Linguistic Notes from La Jolla.

Dalemans, W., Gillis, S. & Durieux, G. (1994). The acquisition of stress: A data-oriented Approach. *Computational Linguistics 3*(20), 420-451.

Dresher, B. & Kaye, J. (1990). A computational learning model for metrical theory. *Cognition* 34, 137-195.

Elman, J.L., Bates, E.A., Johnson, M.H., Karmiloff-Smith, A., Parisi, D. & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development.* Cambridge, MA: MIT Press.

Fikkert, P. (1994). *On the acquisition of prosodic structure*. Ph.D. Thesis, Holland Institute of Generative Linguistics.

Flemming, E. (1995). *Auditory representations in phonology*. Ph.D. Thesis, UCLA.

Gentner, D. (1982).Why Nouns are learned before Verbs: Linguistic relativity versus natural partitioning. In S.A. Kuczaj (Ed.) *Language development Vol. 2: Language Thought and Culture*, (pp. 301-334) .Lawrence Erlbaum Associates.

Goldsmith, J. (1990). *Autosegmental and metrical phonology*. Cambridge MA: Basil Blackwell.

Halle, M. & Mohanan, K. (1985).Segmental phonology of Modern English *Linguistic Inquiry* 16 (1), 57-116.

Hammond, M. (1990). Parameters of metrical theory and learnability. In I. Roca (Ed.), *Logical issues in language acquisition* (pp. 47-62). Dordrecht: Foris.

Hayes, B. *A metrical theory of stress.* Chicago IL: Chicago University Press.

Hayes, B. (1997). Phonetically driven phonology: The role of optimality theory and inductive grounding. *Proceedings of the 1996 Milwaukee conference on formalism and functionalism in linguistics*. Milwaukee WI.

van der Hulst, H. (1984). *Syllable structure and Dutch*. Dordrecht: Foris.

Inkelas, S., Orgun, C. & Zoll, C. (1996). *Exceptions and static phonological patterns: cophonologies vs. prespecification*. (Unpublished Manuscript: U.C. Berkeley and University of Iowa)

Ito, J. & Mester, A. (1986).The phonology of voicing in Japanese. *Linguistic Inquiry* 17, 49-73.

Ito, J. & Mester, A. (1996). *Structural economy and OCP interaction in local domains*. Handout of paper presented at Western Conference on Linguistics University of California, Santa Cruz.

Ito, J., Mester, A. & Padgett, J. (1995). Licensing and underspecification in Optimality Theory, *Linguistic Inquiry* 26(4), 571-613.

Joanisse, M.F. (1999). Exploring syllable structure with connectionist networks. *Proceedings of the XIVth International Congress of Phonetic Sciences.* San Francisco CA.

Joanisse, M.F. & Seidenberg, M.S. (1999). Impairments in verb morphology following brain injury: A connectionist model. *Proceedings of the National Academy of Science, USA* 96(13), 7592-7597.

Kager, R. (1989). *A metrical theory of stress and destressing in English and Dutch*. Dordrecht: Foris.

Liljencrants, J. & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Journal of Phonetics* 48(4), 839-862.

Lindblom, B., MacNeilage, P. & Studert-Kennedy, M. (1984). *Self-organizing processes and the explanation of phonological universals.* In B. Butterworth, B. Comrie & O. Dahl (Eds), *Explanations for Language Universals.* Berlin: Mouton.

Macnamara, J. (1972). Cognitive basis of language learning in infants. *Psychological Review* 79, 1-13.

MacWhinney, B. & Leinbach, J. (1991).Implementations are not conceptualizations: Revising the verb learning model. *Cognition* 29, 121-157.

Marcus, G., Ullman, M., Pinker, S., Hollander, M., Rosen, T. & Xu, F. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development* 57.

Pinker, S. (1991). Rules of language. *Science 253*, 530-535.

Pinker, S. & Prince, A. (1988).On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28(1-2), 73-193.

Prasada, S. & Pinker, S. (1993). Generalisation of regular and irregular morphological patterns. *Language and Cognitive Processes* 8(1), 1-56.

Prince, A. & Smolensky, P. (1993). *Optimality Theory: Constraint interaction in generative grammar*. Rutgers University.

Rumelhart, D.E. & McClelland, J.L. (1986). On learning the past tenses of English verbs. In J.McClelland & D.Rumelhart (Eds), *Parallel distributed processing: Explorations in the microstructure of cognition (vol.2)*. Cambridge, MA: MIT Press.

Sagey, E. (1986). *The representation of features and relations in nonlinear phonology*. Ph.D. Thesis, MIT.

Seidenberg, M.S. & MacDonald, M.C. (in press). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*

Seidenberg, M.S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science* 275, 1599-1603.

Selkirk, E. (1980).Prosodic domains in phonology: Sanskrit revisited. In M. Aronoff & M.-L. Kean (Eds.) *Juncture.* Saratoga, CA: Anma Libri.

Slobin, D. (1985).Crosslinguistic evidence for the language-making capacity. In D. Slobin (Ed.) *The Crosslinguistic Study of Language Acquisition, Vol. 2*. Hillsdale, NJ: Lawrence Erlbaum Associates, pp 1157-1249.

Smith, N. (1973). *The acquisition of phonology: A case study.* Cambridge UK: Cambridge University Press.

Stampe, D. (1979).*A dissertation on natural phonology.* Ph.D. Thesis, University of Chicago.

Steriade, D. (1994). *Positional neutralization and the expression of contrast.* (unpublished manuscript, UCLA)

Stevens, K. N. (1989).On the quantal nature of speech. *Journal of Phonetics* 17, 3-45.

Tesar, B. & Smolensky, P. (1996). *Learnability in Optimality Theory* (TR 96-3). Johns Hopkins University Department of Cognitive Science.

Wijnen, F. (1992). Incidental word and sound errors in young speakers. *Journal of Memory and Language* 31, 734-755.

Williams, R. J. & Peng, J. (1990). An efficient gradient-based algorithm for on-line training of recurrent network trajectories. *Neural Computation* 2, 490-501.

Zipf, G.K. (1935). The psycho-biology of language: An introduction to dynamic philology. Boston, MA: Houghton Mifflin.